

УДК 519.723

**Назар Парадюк, студент, Зіновій Горбань, студент, Богдан Чернецький, студент,  
Мирослава Яворська, к.т.н., доцент**

Тернопільський національний технічний університет імені Івана Пулюя

# **ПРОГРАМНА РЕАЛІЗАЦІЯ ПОБУДОВИ СТИСКАЮЧОГО КОДУ ДЛЯ ОПТИМІЗАЦІЇ ОБРОБКИ ДАНИХ**

Реалізовано принцип побудови коду Шенона-Фано для кодування масиву даних що  
носять стохастичний характер.

Ключові слова: оптимальне кодування, префіксний код, ентропія повідомлення.

**Nazar Paradiuk, Zinovii Horban, Bohdan Chernetskyi, Myroslava Yavorska**  
**SOFTWARE IMPLEMENTATION OF COMPRESSIVE CODE TO DATA**  
**PROSESING OPTIMIZATION**

Shannon-Fano code software implementation for encoding of stochastic data was realized.

Keywords: optimal encoding, prefix code, entropy of message.

Реалізовано програмне забезпечення в середовищі MATLAB для статистичної  
обробки масиву вихідних даних та їх кодування префіксним кодом із стискаючими  
властивостями. Алгоритм використовує коди змінної довжини: символ, який часто  
зустрічається, кодується кодом меншої довжини, а той що рідше зустрічається – кодом  
більшої довжини. Таким чином досягається стискання записуваної в пам'ять  
інформації. Префіксність коду дозволяє однозначно декодувати будь-яку послідовність  
кодів слів без використання розділювачів оскільки жодне коротше кодове слово не  
є префіксом коду більшої довжини. Неоднозначність при побудові коду  
регламентується обмеженням, сформульованим в I теоремі Шенона:

$$l_{сер} \geq -\sum_{i=1}^n p_i \log_2 p_i$$
 де,  $p_i$  – імовірність слідування i-го символу в повідомленні,  $l_{сер}$  –  
середня довжина кодової комбінації.

В табл. 1 приведені отримані при допомозі розробленого програмного  
забезпечення результати статистичного аналізу та коди символів для набору даних [1 3  
7 4 1 0 9 6 7 3 2 8 1 4 3 0 9 7 5 5 9 1 3 5 4 4 2 8 5 4 7 8 9 1 3 4 1 2 1 6 8 5 3 4 7 6 2 3 9 1].

Символи	0	1	2	3	4	5	6	7	8	9
Частота появи	0.16	0.14	0.14	0.10	0.10	0.10	0.08	0.08	0.06	0.04
Код	11	101	100	011	010	0011	0010	0001	00001	00000

Кодований запис даного набору матиме вигляд:

1010110001010101110000000100001011100000011010100111100000001001100110  
000010101100110100101000000100110100001000010000011011010111001100100000100  
11011010000100101000110000011

При ентропії  $H = 3.23$  середня довжина кодової комбінації  $L_{сер} = 3.3$  біт, довжина  
кодованого повідомлення 165 біт тоді як нестиснений запис потребуватиме 200 біт.